35 p.

RIAS

N64-18447

CODE-1

CR-53567

TECHNICAL REPORT

64-3

FEBRUARY 1964

OTS PRICE

XEROX $ 3.60 ph

MICROFILM $ 1.25 mf.

SOME APPLICATIONS OF STOCHASTIC DIFFERENTIAL EQUATIONS TO OPTIMAL NONLINEAR FILTERING

By

W. M. Wonham

SOME APPLICATIONS OF STOCHASTIC DIFFERENTIAL EQUATIONS TO OPTIMAL NONLINEAR FILTERING

by

W. M. Wonham

Feb~~ruary~~. 1964 35p refs

CENTER FOR CONTROL THEORY
Research Institute for Advanced Studies (RIAS)
A Division of the Martin Co.
7212 Bellona Avenue
Baltimore, M~~aryland 21212~~d.

1482033

Martin Co., RIAS Div.

(NASA Contract NASw-845; Contract AF 49(638)-1206)
rept. (NASA CR-53567; RIAS-TR-64-3) OTS: $3.60 ph, $1.25 mf

SOME APPLICATIONS OF STOCHASTIC DIFFERENTIAL EQUATIONS TO OPTIMAL NONLINEAR FILTERING[1]

by

W. M. Wonham

1. Introduction.

A current problem in control theory is that of estimating the dynamical state of a physical system, on the basis of data perturbed by noise. Solution of the estimation problem is usually immediate if one knows the probability distribution of the system state at each instant of time, conditional on the data available up to that instant. It is therefore of interest to ask how this posterior probability distribution evolves with time, and if possible to specify the dynamical structure of a filter (i.e. analog device) which generates the posterior distribution when its input is the time function actually observed.

In the present report, filters of this type are defined by means of stochastic differential equations[2] for the posterior distribution in which the observed time function appears as a forcing term. Differential equations for this purpose were introduced in 1960 by Stratonovich [1], who also indicated their application to stochastic control problems [2]. When the dynamical system under observation is linear and the noise is white Gaussian it has been shown [3] that Stratonovich's equation can be solved formally to yield the stochastic differential equation of the optimal (linear) filter. When the function to be estimated is a Markov step process and the noise is white Gaussian the optimal (nonlinear) filter equations were stated in [4]. The latter equations are discussed in more detail in sect. 3, below; they differ from those of Stratonovich in a sense to be noted in the sequel. For one example, discussed in sect. 3, performance of the optimal nonlinear filter is evaluated numerically and is found to be substantially better than that of the simpler Wiener filter.

---

1. This research was supported in part by the United States Air Force through the Air Force Office of Scientific Research, under Contract No. AF 49(638)-1206, and in part by the National Aeronautics and Space Administration under Contract No. NASw-845.

2. A brief review of stochastic differential equations is given in Appendix 1.

In sect. 4, the equations of sect. 3 are generalized heuristically to the case where the state space of the step process is continuous, and in sect. 5 some tentative remarks are made on the form of the solutions.

Some parallel work on noisy observation of a diffusion process will be reported by Kushner in a forthcoming paper [5].

2. Noisy measurement of an unknown constant.

The basic idea of a 'functional' filter is illustrated by the following simple estimation problem. Let x be a discrete, real-valued random variable with range of values $a_1, \ldots, a_K$ and a priori probability distribution $\{p_j(0),\ j = 1, \ldots, K\}$ at $t = 0$. Suppose that one observes the function

$$y(t) = xt + \int_0^t \beta(s)dw(s), \quad t \geq 0, \tag{1}$$

where the function β is known to the observer,[3] and $\{w(t),\ t \geq 0\}$, is a Wiener process which is independent of x, with $P\{w(0) = 0\} = 1$. The process $y(t)$ defined by (1) can also be written as the solution of the stochastic differential equation

$$dy(t) = x\,dt + \beta(t)dw(t), \quad y(0) = 0. \tag{2}$$

Dividing formally by dt one obtains the possibly more familiar version

$$\dot{y}(t) = x + \beta(t)\dot{w}(t), \quad y(0) = 0, \tag{3}$$

where $\dot{w}$ represents Gaussian white noise. Since w is not differentiable in the ordinary sense we shall use instead the differential notation of (2) and interpret (2) as an equation of Ito's type (Appendix 1).

3. We shall assume that β is continuously differentiable, and bounded away from 0 for $t \geq 0$.

Let us now introduce the posterior distribution

$$p_j(t) = P\{x = a_j | y(s),\quad 0 \leq s \leq t\}, \quad j = 1, \ldots, K. \tag{4}$$

Evaluation of $p_j(t)$ is straightforward (see Appendix 2); the result is[4]

$$p_j(t) = \frac{p_j(0)\exp[a_j \int_0^t \beta(s)^{-2} dy(s) - \frac{1}{2} a_j^2 \int_0^t \beta(s)^{-2} ds]}{\sum_{k=1}^{K} p_k(0)\exp[a_k \int_0^t \beta(s)^{-2} dy(s) - \frac{1}{2} a_k^2 \int_0^t \beta(s)^{-2} ds]}. \tag{5}$$

The stochastic integral in (5) is well-defined ([6] IX,§2). Now write $p(t) = [p_1(t), \ldots, p_K(t)]$ and consider the joint process $\{x, p(t), t \geq 0\}$, where we regard x as a fixed random variable with distribution $\{p_j(0),\ j=1, \ldots, K\}$. Since almost every $w(t)$ sample function is continuous the same is true of the $p(t)$ sample functions (by [6] IX, Thm. 5.2). Moreover, it is easily seen (Appendix 3) that the $\{x, p(t)\}$ process is Markov.

Our aim is to describe the evolution in time of the p_j's by means of a system of stochastic differential equations. Having verified the existence of the limits (8) and (9) written below, one can apply to the Markov process $\{x, p(t)\}$ a representation theorem of Doob ([6] VI, Thm. 3.3). Alternatively, since the $p_j(t)$ are known explicitly, it is more direct to apply a result of Dynkin (Appendix 1 or [7], Thm. 7.2), and this gives

$$dp_j(t) = m_j[t, x, p(t)]dt + \sigma_j[t, x, p(t)]dw(t), \quad j = 1, \ldots, K. \tag{6}$$

The functions m_j and σ_j in (6) have the interpretation

$$m_j(t, \xi, \pi) = \lim_{h \to 0} E\{\frac{p_j(t + h) - p_j(t)}{h} \Big| x = \xi, p(t) = \pi\} \tag{7}$$

and

4. The qualification 'with probability 1' on equalities between conditional probabilities is to be understood.

$$\sigma_i(t, \xi, \pi)\sigma_j(t, \xi, \pi) = \lim_{h \to 0} E\left\{\frac{[p_i(t+h) - p_i(t)][p_j(t+h) - p_j(t)]}{h}\right| \tag{8}$$

$$x = \xi,\ p(t) = \pi\};$$

$$i, j = 1, \ldots, K.$$

In Appendix 3 the limits (7) and (8) are computed from (5), using Dynkin's formulas; the results are

$$m_j(t, x, p) = \beta(t)^{-2}(x - \bar{x})(a_j - \bar{x})p_j \tag{9}$$

$$\sigma_j(t, x, p) = \beta(t)^{-1}(a_j - \bar{x})p_j \tag{10}$$

$(j = 1, \ldots, K)$, where

$$\bar{x} = \sum_{k=1}^{K} a_k p_k. \tag{11}$$

Writing out (6) in full and noting (2), we obtain finally

$$dp_j(t) = - \beta(t)^{-2}\bar{x}(t)[a_j - \bar{x}(t)]p_j(t)dt + \\ + \beta(t)^{-2}[a_j - \bar{x}(t)]p_j(t)dy(t), \tag{12}$$

$$j = 1, \ldots, K.$$

The system of stochastic differential equations (12) is the desired result. It can be interpreted as specifying the structure of a filter (or ideal analog device) which continuously generates the posterior distribution $p(t)$ when the input is the observed function $y(t)$. From a practical viewpoint this interpretation might be useful when the function actually observed is not y, but rather some approximation to the 'function' $\dot{y}$ indicated in (3). In that case formal division by dt in (12) yields a system of nonlinear differential equations for the p_j's, in which $\dot{y}$ appears as a forcing term. This system of equations defines the filter.

Example. It is worth emphasizing that the ordinary rules of integration cannot be applied to the stochastic equation (12) in an attempt to regain the explicit solution (5). To illustrate this fact let x have possible values $a_1 = +1$, $a_2 = -1$ and let $p_1(0) = p_2(0) = \frac{1}{2}$, $\beta \equiv 1$. Since $p_1(t) + p_2(t) \equiv 1$ it is sufficient to consider

$$q(t) = p_1(t) - p_2(t). \tag{13}$$

From (11), $\bar{x}(t) = q(t)$; and from (12) the filter is defined by

$$dq = -q(1 - q^2)dt + (1 - q^2)dy. \tag{14}$$

On the other hand (4) yields the evaluation

$$q(t) = \tanh [y(t)]; \tag{15}$$

and formal differentiation of (15) gives

$$dq = (1 - q^2)dy. \tag{16}$$

We observe that the first term on the right side of (14) is absent from (16). The point is not only that stochastic differential equations of Ito type cannot be manipulated by the usual formal rules (cf. [6] IX, §5) but also that an analog device for generating q should be set up according to the Ito equation (14), and not according to the 'formal' equation (16).[5]

It must be noted finally that Stratonovich's procedure [1] applied to this example leads to (16) and not to (14) (cf. [2], eq. (9)).

---

5. The Ito equation (14) is solved (Appendix 1) by replacing (14) by an integral equation and constructing the solution q by successive approximations. The structure of the analog feedback device should be chosen to model that of the successive approximation scheme. Whether or not physical analog devices can actually be built to operate in this way on 'white' noise (i.e. wideband random inputs) is a question open to investigation.

3. Noisy observations of a Markov step process.

Let $\{x(t), t \geq 0\}$, be a stationary Markov step process with a finite number of states (step levels) $a_1, \ldots, a_K$.[6] Denote the transition probabilities by

$$p_{ij}(h) = P\{x(t+h) = a_j | x(t) = a_i\}.$$

We assume that

$$p_{ij}(h) = \begin{cases} 1 - \nu_i h + o(h), & j = i \quad (h \to 0) \\ \nu_{ij} h + o(h), & j \neq i \quad (h \to 0) \end{cases} \tag{17}$$

where the $\nu_{ij} \geq 0$ are constants and

$$\nu_i = \sum_{\substack{j=1 \\ j \neq i}}^{K} \nu_{ij}, \quad i = 1, \ldots, K. \tag{18}$$

Let the distribution of $x(0)$ be $\{p_j(0), \; j = 1, \ldots, K\}$.

As in sect. 2, suppose that the process observed is $\{y(t), t \geq 0\}$ defined by

$$dy(t) = x(t)dt + \beta(t)dw(t), \quad t \geq 0, \tag{19}$$

where $P\{y(0) = 0\} = 1$ and the Wiener process $\{w(t), \; t \geq 0\}$ is independent of the $x(t)$ process. Introduce the posterior probabilities

$$p_j(t) = P\{x(t) = a_j | y(s), \quad 0 \leq s \leq t\}, \tag{20}$$

$$j = 1, \ldots, K.$$

As before, we seek a stochastic differential equation for the $p_j(t)$. The appropriate generalization of (12) is obtained in Appendix 4; the result is

6. For a complete discussion see [6] VI, §1, where our ν_i, ν_{ij} are denoted by q_i, q_{ij}. The $x(t)$ process can be defined so as to be separable and measurable; integrals such as $\int_0^t x(s)ds$ are then well-defined random variables.

$$dp_j(t) = [-\nu_j p_j(t) + \sum_{\substack{i=1\\ i\neq 1}}^{K} \nu_{ij} p_i(t)]dt -$$

$$- \beta(t)^{-2}\bar{x}(t)[a_j - \bar{x}(t)]p_j(t)dt + \tag{21}$$

$$+ \beta(t)^{-2}[a_j - \bar{x}(t)]p_j(t)dy(t), \quad j = 1, \ldots, K.$$

Eq. (21) can be interpreted as specifying the structure of an ideal analog device for generating the p_j's from the data, y. Comparison of (21) with (12) shows that the only novel feature of (21) is the first term on the right side. This term is of the form $\mathcal{L}^+[p]dt$ where $\mathcal{L}^+$ is the forward operator of the $x(t)$ process, and thus represents a change in p due to the observer's a priori knowledge of how the $x(t)$ process evolves.

We shall now discuss a simple special case of (21) in detail.

Example. Let $\beta \equiv$ const. and suppose that x is a 'random telegraph signal' [11]; that is,

$$a_1 = +1, \quad a_2 = -1 \tag{22}$$

$$\nu_i = \nu_{ij} = \nu, \quad i, j = 1, 2.$$

The parameter ν is the expected number of jumps of $x(\cdot)$ in unit time. Let

$$q(t) = p_1(t) - p_2(t). \tag{23}$$

From (21) and (22),

$$dq = -2\nu q dt - \beta^{-2}q(1 - q^2)dt + \beta^{-2}(1 - q^2)dy \tag{24}$$

or equivalently

$$dq = [-2\nu q - \beta^{-2}(1 - q^2)(q - x)]dt + \beta^{-1}(1 - q^2)dw. \tag{25}$$

We shall evaluate the filter performance in terms of mean square estimation error. Thus the optimal estimate of $x(t)$ is

$$\begin{aligned}\hat{x}(t) &= E\{x(t)\,|\,y(s),\quad 0 \leq s \leq t\} \\ &= a_1 p_1(t) + a_2 p_2(t) \\ &= q(t).\end{aligned} \tag{26}$$

Now consider the joint Markov process $\{x(t),\ q(t),\ t \geq 0\}$. It will be assumed that this process has stationary densities $\pi^{\pm}(q)$, $-1 \leq q \leq 1$, defined by

$$\pi^{\pm}(q)dq = P\{x(t) = \pm 1,\quad q(t) \in (q,\ q + dq)\}. \tag{27}$$

Then the mean square estimation error is

$$\sigma^2 = \int_{-1}^{1} (1 - q)^2 \pi^{+}(q)dq + \int_{-1}^{1} (1 + q)^2 \pi^{-}(q)dq. \tag{28}$$

It remains to calculate the densities $\pi^{\pm}$. By inspection of (25), the stationary Kolmogorov equation of the $\{x,\ q\}$ process is

$$\begin{aligned}&\frac{1}{2}\,\beta^{-2}[(1 - q^2)^2 \pi^{\pm}(q)]'' - \\ &\quad - [\beta^{-2}(\pm 1 - q)(1 - q^2)\pi^{\pm}(q) - 2\nu q\ \pi^{\pm}(q)]' \pm \\ &\qquad \pm \nu[\pi^{-}(q) - \pi^{+}(q)] = 0\end{aligned} \tag{29}$$

where $(')$ denotes d/dq. With the symmetry condition $\pi^{-}(q) = \pi^{+}(-q)$, (29) has the unique solution

$$\pi^{\pm}(q) = c(1 \mp q)^{-1}\exp[-\,2\mu(1 - q^2)^{-1}] \tag{30}$$

where

$$c = [2\int_{1}^{\infty} z^{1/2}(z - 1)^{-1/2}e^{-2\mu z}dz]^{-1} \tag{31}$$

and

$$\mu = \beta^2 \nu. \tag{32}$$

From (28), (30) the stationary error variance is then

$$\sigma^2 = \frac{\int_0^\infty z^{-1/2}(z+1)^{-3/2} e^{-2\mu z} dz}{\int_0^\infty z^{-1/2}(z+1)^{1/2} e^{-2\mu z} dz}. \tag{33}$$

It is easy to show that $\sigma^2 \to 0$ as $\mu \to 0$, and that

$$\sigma^2 = 1 - (2\mu)^{-1} + O(\mu^{-2}) \quad \text{as} \quad \mu \to \infty. \tag{34}$$

The result (33) will be compared with the error variance of a Wiener filter which is optimal for the same input. For the Wiener filter a standard computation yields

$$\begin{aligned} \sigma_W^2 &= 2\mu[(1 + \mu^{-1})^{1/2} - 1] \\ &= 1 - (4\mu)^{-1} + O(\mu^{-2}) \quad \text{as} \quad \mu \to \infty. \end{aligned} \tag{35}$$

Numerical results are given in Fig. 1. Since the nonlinear filter generates the estimate defined by (26) it is necessarily optimal (with respect to error variance) in the class of all filters which operate on the present and past of the data y. Thus $\sigma^2 \leq \sigma_W^2$ and in fact σ^2 is substantially less than σ_W^2 except when the noise level is very high.

4. Generalization to continuous state space.

The differential equations (12), (21) were derived on the assumption that the state space of the $x(t)$ process is a finite set. The following is a heuristic generalization to a continuous state space. Let $\{x(t),\ t \geq 0\}$ be a real-valued Markov step process with state space X, where X is a closed finite interval. Let $\nu(\xi)$, $\nu(\xi, A)$ be defined for ξ in X and A a Borel subset of X; the functions $\nu(\cdot)$, $\nu(\cdot,\cdot)$ are assumed to be a 'standard pair' in the sense of Doob [6] VI, §2.[6] In addition, $\nu(\cdot)$ is assumed to be bounded on X; the $x(t)$ sample functions are then almost all step functions [6]. In analogy to (17) one has

$$p(h, \xi, A) = P\{x(t+h) \in A | x(t) = \xi\}$$

$$= \begin{cases} \nu(\xi, A)h + o(h), & \xi \notin A \\ 1 - \nu(\xi)h + o(h), & A = \{\xi\}. \end{cases} \tag{36}$$

As in sect. 3, suppose next that

$$dy(t) = x(t)dt + \beta(t)dw(t), \quad t \geqq 0, \tag{37}$$

and introduce the posterior probability

$$p(t, A) = P\{x(t) \in A | y(s), \quad 0 \leqq s \leqq t\}. \tag{38}$$

Then inspection of (21) suggests the generalization

$$\begin{aligned} dp(t, A) = & [- \int_A \nu(\xi, X - A)p(t, d\xi) + \\ & + \int_{X-A} \nu(\xi, A)p(t, d\xi)]dt - \\ & -[\beta(t)^{-2}\bar{x}(t) \int_A [\xi - \bar{x}(t)]p(t, d\xi)]dt + \\ & +[\beta(t)^{-2} \int_A [\xi - \bar{x}(t)]p(t, d\xi)]dy(t). \end{aligned} \tag{39}$$

In (39),

$$\bar{x}(t) = \int_X \xi p(t, d\xi), \tag{40}$$

6. The functions ν are denoted in [6] by q.

and A is an arbitrary Borel subset of X.

Just as in the case of a (finite-dimensional) Ito equation, (39) might plausibly be interpreted by starting from the corresponding integral equation, obtained by integrating both sides of (39) with respect to t, and defining the solution as the limit of successive approximations. Unlike (12) and (21), however, the general equation (39) cannot readily be interpreted as specifying the dynamics of a practically realizable filter for generating p from the data y.

5. Sufficient statistics.

We have seen in sect. 4 that the stochastic differential equation for the posterior distribution of $x(t)$ cannot be used directly, in general, to design an optimal filter. In practice, construction of the posterior distribution must be reduced to the evaluation of a 'small' number of functionals (sufficient statistics) on the observed function y. For example, inspection of (5) shows that the K-dimensional stochastic system (12) can be replaced by the 1-dimensional system

$$d\sigma(t) = \beta(t)^{-2} dy(t), \quad \sigma(0) = 0. \tag{41}$$

which generates the sufficient statistic

$$\sigma(t) = \int_0^t \beta(s)^{-2} dy(s). \tag{42}$$

On the other hand the writer knows of no similar reduction of the system (21) or of (39).

Even if a nontrivial sufficient statistic $z(t)$ for the determination of p exists, it may be impossible to write z as an explicit functional of y. It would be enough to know, however, that z satisfies a stochastic differential equation

$$dz = \zeta_1(t, z)dt + \zeta_2(t, z)dy \qquad (43)$$

where the functions ζ_1, ζ_2 are known; then in principle z could be obtained as the output of an analog device set up according to (43). Thus the 'solution' of an equation of type (21), (39) might take the form of a (known) function of a statistic z which satisfies a (known) equation of type (43). The investigation of solutions of this type (if they exist) would be of considerable interest.

Appendix 1. Stochastic differential equations.

1. Since stochastic differential equations are not yet widely used in engineering applications we summarize here some definitions and known results. For a detailed account the reader is referred to Doob [6] VI, §3, and Dynkin [7], Chs. 7 and 11.

Let $\{z(t),\ t \geqq 0\}$ be a stochastic process in K-dimensional Euclidean space R^K; we write $z = (z_1, \ldots, z_K)$, where the z_i are real-valued, and put $\|z\| = (\sum_1^K z_i^2)^{1/2}$. Let $\{w(t),\ t \geqq 0\}$ be a Wiener process (Brownian motion process) in R^J; that is, $w(t) = [w_1(t), \ldots, w_J(t)]$ where the $w_i(t)$ are independent Wiener processes in R^1 and, for $t \geqq s \geqq 0$,

$$
\begin{aligned}
E\{[w_i(t) - w_i(s)][w_j(t) - w_j(s)]\} &= t - s, \quad j = i; \\
&= 0, \qquad\quad j \neq i.
\end{aligned}
\tag{44}
$$

(See [6] II, §9 for the definition of the Wiener process in R^1).

The stochastic differential equation of interest here is written

$$
dz(t) = m[t, z(t)]dt + \sigma[t, z(t)]dw(t), \quad t \geqq 0, \tag{45}
$$

where m is a K-vector and σ is a $K \times J$ matrix. Loosely interpreted, (45) states that in a small time interval $(t, t + dt)$ the vector $z(t)$ suffers a 'dynamical' displacement $m[t, z(t)]dt$ plus a random displacement $\sigma[t, z(t)]dw(t)$, where the latter is a Gaussian random vector with mean 0 and covariance matrix $\sigma[t, z(t)]\sigma'[t, z(t)]dt$ (' denotes transpose).

Dividing both sides of (45) formally by dt one obtains

$$
\dot{z} = m(t, z) + \sigma(t, z)\dot{w} \quad (\cdot = d/dt) \tag{46}
$$

where $\dot{w}$ is a J-vector whose components are independent 'Gaussian white noise processes'. The notation of (46) has been more common in the engineering literature than that of (45) but it is objectionable for two reasons:

(a) Almost all $w(t)$ sample functions are nondifferentiable almost everywhere, a fact which is intuitively plausible if we note that $E\|dw(t)\|$ is proportional to $(dt)^{\frac{1}{2}}$.

(b) When applied to (46) the usual formal rules of integration lead in general to results which are definitely incorrect. This statement will be illustrated later.

2. The stochastic differential equation (45) does not specify the value of a derivative. The equation can be defined by giving an explicit procedure for constructing the solution, and then has meaning only insofar as this construction can be carried out (cf. Gihman [9]). An alternative interpretation of (45) due to Ito [8] is the following. Replace (45) by the stochastic integral equation

$$z(t) = z(0) + \int_0^t m[s, z(s)]ds + \int_0^t \sigma[s, z(s)]dw(s). \tag{47}$$

The stochastic integrals on the right side of (47) are defined ([6] IX, §2, §5) under suitable restrictions on the (random) functions m and σ. Ito's construction of a $z(t)$ process which satisfies (47) is carried out by successive approximation; one sets $z^{(o)}(t) \equiv 0$ and

$$z^{(n+1)}(t) = z(0) + \int_0^t m[s, z^{(n)}(s)]ds + \int_0^t \sigma[s, z^{(n)}(s)]dw(s), \quad n = 0,1,2, \dots . \tag{48}$$

Conditions under which the sequence $\{z^{(n)}\}$ converges for t in a finite interval $[0, T]$ are given by Doob ([6] IX, §3)[7] and are, mainly, that the

7. Doob's treatment for $K = J = 1$ can be generalized after replacing $|m|$ by $\|m\|$ and $|\sigma|$ by $\|\sigma\| = (\sum_i \sum_j \sigma_{ij}^2)^{\frac{1}{2}}$ (cf. [7], Ch. 7).

functions $m(t, z)$, $\sigma(t, z)$ satisfy a uniform Lipschitz condition in z, and are bounded in norm by $C(1 + \|z\|^2)^{\frac{1}{2}}$ where C is some constant. Then there exists a process $\{z(t),\ 0 \leq t \leq T\}$ with the following properties:

(i) $\lim_{n \to \infty} z^{(n)}(t) = z(t)$ uniformly in t with probability 1, and the $z(t)$ process is essentially unique.

(ii) The $z(t)$ sample functions are almost all continuous in $[0, T]$.

(iii) For each $t \in [0, T]$, (47) is true with probability 1.

(iv) If the initial value $z(0)$ is a random variable which is independent of the increments $\{w(t_2) - w(t_1),\ t_1, t_2 \in [0, T]\}$ then $\{z(t),\ 0 \leq t \leq T\}$ is a Markov process.

In addition, the $z(t)$ process has the following local properties, which make precise the interpretation of (45) given earlier:

$$\text{(v)} \quad \lim_{h \to 0} E\left\{\frac{z(t+h) - z(t)}{h} \,\middle|\, z(t) = \zeta\right\} = m(t, \zeta) \tag{49}$$

$$\text{(vi)} \quad \lim_{h \to 0} E\left\{\frac{[z_i(t+h) - z_i(t)][z_j(t+h) - z_j(t)]}{h} \,\middle|\, z(t) = \zeta\right\}$$

$$= \sum_{r=1}^{J} \sigma_{ir}(t,\zeta)\sigma_{jr}(t,\zeta) \tag{50}$$

$$= b_{ij}(t,\zeta), \text{ say}; \quad i,j = 1, \ldots, K.$$

3. Assume for the moment that the Markov process $\{z(t),\ 0 \leq t \leq T\}$ constructed above has a transition probability density $p = p(s, z; t, \zeta)$, defined for $0 \leq s < t \leq T$ and $z, \zeta \in R^K$. It would be convenient if the hypotheses made on m and σ (strengthened to include the differentiability needed below) guaranteed that the density p exists and satisfies the Kolmogorov equations

$$-\frac{\partial p}{\partial s} = \frac{1}{2}\sum_{i=1}^{K}\sum_{j=1}^{K} b_{ij}(s, z)\frac{\partial^2 p}{\partial z_i \partial z_j} + \sum_{i=1}^{K} m_i(s, z)\frac{\partial p}{\partial z_i} \tag{51}$$

$$\frac{\partial p}{\partial t} = \frac{1}{2}\sum_{i=1}^{K}\sum_{j=1}^{K}\frac{\partial^2}{\partial \zeta_i \partial \zeta_j}[b_{ij}(t,\zeta)p] - \sum_{i=1}^{K}\frac{\partial}{\partial \zeta_i}[m_i(t,\zeta)p] \tag{52}$$

Unfortunately it does not seem possible to establish (51) and (52) without making a priori assumptions on the smoothness of p (see the discussion in [6] VI, §3). Nevertheless, if the $z(t)$ process obtained by solving the Ito equation (45) has a transition density which satisfies the Kolmogorov equations, then the coefficients which appear in the latter are related to the functions m and σ of the Ito equation according to (49) - (52) above. Used heuristically, this correspondence between (45) and (51), (52) is convenient in engineering applications (see e.g. [10]).

4. The following example[8] shows that Ito equations cannot be manipulated by the ordinary rules of integration (cf. also [6], p. 443). Let $\{w(t), t \geqq 0\}$ be a Wiener process in R^1 with $P\{w(0) = 0\} = 1$. Let

$$z(t) = e^{w(t)}, \quad t \geqq 0. \tag{53}$$

It can be shown that the $z(t)$ process can be represented as the solution of an Ito equation (45). From (49), (50) we find

$$\begin{aligned} m(\zeta) &= \frac{1}{2}\zeta \\ \sigma(\zeta) &= \zeta \end{aligned} \tag{54}$$

so that

$$dz(t) = \frac{1}{2} z(t)dt + z(t)dw(t), \quad t \geqq 0, \tag{55}$$

8. Suggested to the writer by H. J. Kushner.

with initial condition $P\{z(0) = 1\} = 1$. On the other hand if (55) is replaced by

$$\dot{z} = \frac{1}{2}z + z\dot{w}, \quad z(0) = 1, \tag{56}$$

and the last equation integrated formally, the result is

$$z(t) = e^{\frac{1}{2}t + w(t)}. \tag{57}$$

5. It is useful to know under what conditions a given process $\{\psi(t),\ t \geq 0\}$ can be represented as the solution of an Ito equation (45) or equivalently of the integral equation (47). One such representation theorem is given by Doob ([6] VI, Thm. 3.3). We will state here a slightly specialized version of a theorem of Dynkin ([7], Thm. 7.2).

<u>Theorem.</u> <u>Let the process</u> $\{z(t),\ t \geq 0\}$ <u>satisfy an integral equation of form</u> (47) [<u>in particular we can have</u> $z(t) \equiv w(t)$] <u>and let</u> $\varphi = \varphi(t, \zeta)$ <u>be a numerical function, twice continuously differentiable in</u> (t, ζ) <u>for</u> $t \geq 0$ <u>and</u> ζ <u>in</u> R^K. <u>Put</u> $\psi(t) \equiv \varphi[t, z(t)]$. <u>Then the process</u> $\{\psi(t),\ t \geq 0\}$ <u>satisfies the integral equation</u>

$$\begin{aligned}\psi(t) = \psi(0) &+ \int_0^t \tilde{m}[s, z(s)]ds + \\ &+ \int_0^t \sum_{r=1}^{J} \tilde{\sigma}_r[s, z(s)]dw_r(s).\end{aligned} \tag{58}$$

<u>The functions</u> $\tilde{m}$ <u>and</u> $\tilde{\sigma}_r$ <u>are given by</u>

$$\begin{aligned}\tilde{m}(t, \zeta) = \frac{\partial\varphi(t, \zeta)}{\partial t} &+ \sum_{i=1}^{K} \frac{\partial\varphi(t, \zeta)}{\partial\zeta_i} m_i(t, \zeta) + \\ &+ \frac{1}{2}\sum_{i=1}^{K}\sum_{j=1}^{K} \frac{\partial^2\varphi(t, \zeta)}{\partial\zeta_i\,\partial\zeta_j}\, b_{ij}(t, \zeta),\end{aligned} \tag{59}$$

<u>where</u>

$$b_{ij}(t, \zeta) = \sum_{r=1}^{J} \sigma_{ir}(t, \zeta)\sigma_{jr}(t, \zeta); \tag{60}$$

and

$$\tilde{\sigma}_r(t, \zeta) = \sum_{i=1}^{K} \frac{\partial\varphi(t, \zeta)}{\partial\zeta_i} \sigma_{ir}(t, \zeta), \quad r = 1, \ldots, J. \tag{61}$$

Eq. (58) is (by definition) equivalent to the stochastic differential equation

$$d\psi(t) = \tilde{m}[t, z(t)]\, dt + \sum_{r=1}^{J} \tilde{\sigma}_r[t, z(t)]dw_r(t). \tag{62}$$

Eq. (62) is more general than (45) in the sense that $\tilde{m}$ and the $\tilde{\sigma}_r$ may not be expressible as functions of (t, ψ); however, by adjoining (62) to (45) we obtain a new system of the same type as before.

Finally, it is seen that (62) can also be written

$$d\psi(t) = \tilde{m}[t, z(t)]\, dt + \sum_{i=1}^{K} \frac{\partial\varphi[t, z(t)]}{\partial\zeta_i} dz_i(t) \tag{62'}$$

which exhibits the 'chain rule' explicitly.

Appendix 2. Derivation of (5).

Let $t > 0$ be fixed and put $s_{rn} = rt/n$, $r = 0,1, \ldots, n$. It will be verified that $\hat{p}_j(t)$, defined by

$$\hat{p}_j(t) = \underset{n \to \infty}{\text{l.i.m.}}\ P\{x = a_j | y(s_{rn}),\ \ r = 0,1, \ldots, n\}, \tag{63}$$

is given by the expression on the right side of (5); and then that $\hat{p}_j(t)$ is actually the conditional probability $p_j(t)$ defined by (4). Put $\eta_{rn} = y(s_{rn}) - y(s_{r-1,n})$, $(r = 1, \ldots, n;\ \ n = 1,2, \ldots)$. Then from (1)

$$\eta_{rn} = xt/n + \int_{(r-1)t/n}^{rt/n} \beta(s)dw(s). \tag{64}$$

Thus for each n the random variables $\eta_{rn} - xt/n$, $r = 1, \ldots, n$, are independent and Gaussian with mean 0 and variance

$$v_{rn} = \int_{(r-1)t/n}^{rt/n} \beta(s)^2 ds. \tag{65}$$

Defining

$$\hat{p}_j^{(n)}(t) = P\{x = a_j | y(s_{rn}),\ \ r = 0,1, \ldots, n\}, \tag{66}$$

we have

$$\hat{p}_j^{(n)}(t) = P\{x = a_j | \eta_{rn},\ \ r = 1, \ldots, n\}$$

$$= \frac{p_j(0)\exp[-\sum_{r=1}^{n} (\eta_{rn} - a_j t/n)^2 (2v_{rn})^{-1}]}{\sum_{k=1}^{K} p_k(0)\exp[-\sum_{r=1}^{n} (\eta_{rn} - a_k t/n)^2 (2v_{rn})^{-1}]}$$

$$= \frac{p_j(0)\exp[a_j \sum_{r=1}^{n} \eta_{rn} (rt/n) v_{rn}^{-1} - a_j^2 \sum_{r=1}^{n} (t/n)^2 (2v_{rn})^{-1}]}{\sum_{k=1} p_k(0)\exp[a_k \sum_{r=1}^{n} \eta_{rn} (rt/n) v_{rn}^{-1} - a_k^2 \sum_{r=1}^{n} (t/n)^2 (2v_{rn})^{-1}]}. \tag{67}$$

By definition of the stochastic integral ([6] IX, §2) and the continuity of $\beta(s)^{-2}$, $0 \leqq s \leq t$, there follows

$$\underset{n\to\infty}{\mathrm{l.i.m.}} \sum_{r=1}^{n} \eta_{rn}(rt/n)v_{rn}^{-1} = \int_0^t \beta(s)^{-2}dy(s) \tag{68}$$

and

$$\lim_{n\to\infty} \sum_{r=1}^{n} (t/n)^2\, v_{rn}^{-1} = \int_0^t \beta(s)^{-2}ds. \tag{69}$$

From (67)-(69) it follows that $\hat{p}_j(t)$ coincides with the expression given by (5). It is clear tnat the same result is obtained with any sequence $\{s_{rn}\}$ such that $0 = s_{on} < s_{1n} < \dots < s_{nn} = t$ $(n = 1,2, \dots)$, and $\max_{1 \leqq r \leqq n} (s_{rn} - s_{r-1,n}) \to 0$ as $n \to \infty$.

Let $\mathcal{F}_t$ be the smallest Borel field with respect to which the random variables $y(s)$, $0 \leq s \leq t$, are measurable. To see that $\hat{p}_j(t) = p_j(t)$ note first that $\hat{p}_j(t)$ is certainly measurable relative to $\mathcal{F}_t$. Now suppose that $0 \leq s_\nu \leq t$ and that A_ν is a Borel subset of R^1, $\nu = 1, \dots, N$. If Λ is the event $[y(s_\nu) \in A_\nu,\ \nu = 1, \dots, N]$ then by adjoining the s_ν's to each of the sets $(s_{1n}, \dots, s_{nn})$ $(n = 1,2, \dots)$, and including the $y(s_\nu)$ as conditioning variables in (66), we can write

$$\int_\Lambda \hat{p}_j^{(n)}(t)dP = P\{\Lambda[x = a_j]\},\ n = 1,2, \dots . \tag{70}$$

Since $\hat{p}_j^{(n)} \to \hat{p}_j(t)$ in the mean we obtain, on letting $n \to \infty$,

$$\int_\Lambda \hat{p}_j(t)dP = P\{\Lambda[x = a_j]\}. \tag{71}$$

Since (71) holds for every Λ of the form described, it holds for every Λ in the Borel field $\mathcal{F}_t$ (cf. [6] I, §7). That is, we have verified that $\hat{p}_j(t)$ has the defining properties of the conditional probability $p_j(t)$.

Appendix 3. Properties of the $p(t)$ process (5).

A simple computation from (5) shows that, for $0 < \tau < t$,

$$p_j(t) = \frac{p_j(\tau)\exp[a_j \int_\tau^t \beta(s)^{-2}dy(s) - \frac{1}{2}a_j^2 \int_\tau^t \beta(s)^{-2}ds]}{\sum_{k=1}^{K} p_k(\tau)\exp[a_k \int_\tau^t \beta(s)^{-2}dy(s) - \frac{1}{2}a_k^2 \int_\tau^t \beta(s)^{-2}ds]}. \tag{72}$$

Consider the joint process $\{x, p(t), t \geq 0\}$ where x is regarded as a fixed random variable with distribution $\{p_j(0)\}$. From (72), the vector $p(t)$ depends only on x, $p(\tau)$, and the $w(s)$ increments for $\tau < s < t$. The latter increments are independent of $p(\tau)$, and of x and the $w(s)$ increments for $0 < s < \tau$, on which $p(\tau)$ depends. It follows that the conditional distribution of x, $p(t)$ given x, $p(s)$, $0 \leq s \leq \tau$, is a function of x, $p(\tau)$ alone; that is, the process $\{x, p(t), t \geq 0\}$ is Markov.

The stochastic differential equation (7) for the $p(t)$ process can be established by applying either a representation theorem of Doob([6] VI, Thm. 3.3) or a related theorem of Dynkin ([7], Thm. 7.2). In either case the theorem mentioned must be extended slightly to take account of the fact that only the $p(t)$-component of the joint $\{x, p(t)\}$ process is of diffusion type (alternatively the constant component x can be regarded as a trivial diffusion process). We shall apply Dynkin's theorem, extended to the present case. From (5) we see that $p_j(t)$ is of the form

$$p_j(t) = \phi_j[t, z(t)] \tag{73}$$

where

$$\begin{aligned} z(t) &= \int_0^t \beta(s)^{-2}dy(s) \\ &= \int_0^t x\beta(s)^{-2}ds + \int_0^t \beta(s)^{-1}dw(s) \end{aligned} \tag{74}$$

and

$$\phi_j(t, z) = \frac{p_j(0)\exp[a_j z - \frac{1}{2} a_j^2 \int_0^t \beta(s)^{-2} ds]}{\sum_{k=1}^{K} p_k(0)\exp[a_k z - \frac{1}{2} a_k^2 \int_0^t \beta(s)^{-2} ds]}. \tag{75}$$

Since $\phi_j(t, z)$ is twice continuously differentiable in (t, z), there follows (Appendix 1 or [7], Thm. 7.2)

$$p_j(t) - p_j(0) = \int_0^t m_j[s, x, p(s)]ds + \int_0^t \sigma_j[s, x, p(s)]dw(s). \tag{76}$$

The functions m_j, σ_j are given by

$$m_j(t, x, p) = \frac{\partial}{\partial t} \phi_j(t, z) + x\beta(t)^{-2} \frac{\partial}{\partial z} \phi_j(t, z) + \frac{1}{2} \beta(t)^{-2} \frac{\partial^2}{\partial z^2} \phi_j(t, z) \tag{77}$$

and

$$\sigma_j(t, x, p) = \beta(t)^{-1} \frac{\partial}{\partial z} \phi_j(t, z). \tag{78}$$

The probabilistic meaning of m_j, σ_j is expressed by (7) and (8). The expressions (9) and (10) are computed directly from (75), (77) and (78). Finally, the stochastic differential equation (12) is equivalent (by definition) to the integral equation (76).

Appendix 4. Derivation of (21).

1. We first evaluate $p_j(t)$. To simplify the writing of certain conditional expectations it is convenient to adjoin to the probability space of the $\{x(t), w(t)\}$ process, a 'dummy' step process $\{\tilde{x}(t), t \geq 0\}$, defined to have the same range, initial distribution and transition probabilities as the $x(t)$ process, but independent of $\{x(t)\}$ and $\{w(t)\}$.

Now let $s_{rn} = rt/n$ $(r = 0,1, \ldots, n;\ n = 1,2, \ldots)$ and put

$$\begin{aligned} \eta_{rn} &= y(s_{rn}) - y(s_{r-1,\,n}) \\ \xi_{rn} &= \int_{(r-1)t/n}^{rt/n} x(s)ds \\ \tilde{\xi}_{rn} &= \int_{(r-1)t/n}^{rt/n} \tilde{x}(s)ds. \end{aligned} \tag{79}$$

By (19),

$$\eta_{rn} = \xi_{rn} + \int_{(r-1)t/n}^{rt/n} \beta(s)dw(s);$$

and for each fixed n the random variables $\eta_{rn} - \xi_{rn}$ $(r = 1, \ldots, n)$ are independent and Gaussian with mean 0 and variance

$$v_{rn} = \int_{(r-1)t/n}^{rt/n} \beta(s)^2 ds. \tag{80}$$

Using this fact we can write

$$\begin{aligned} p_j^{(n)}(t) &\overset{\text{df}}{=} P\{x(t) = a_j \mid y(s_{rn}),\quad r = 0,1, \ldots, n\} \\ &= P\{x(t) = a_j \mid \eta_{rn},\quad r = 1, \ldots, n\} \end{aligned} \tag{81}$$

$$= \frac{\sum_{i=1}^{K} p_i(0)p_{ij}(t)E\{\exp[-\sum_{r=1}^{n}(c_{rn} - \xi_{rn})^2(2v_{rn})^{-1}] \mid x_o=a_i,\ x_t=a_j\}_{c_{rn}=\eta_{rn}}}{\sum_{k=1}^{K}\sum_{i=1}^{K} p_i(0)p_{ik}(t)E\{\exp[-\sum_{r=1}^{n}(c_{rn} - \xi_{rn})^2(2v_{rn})^{-1}] \mid x_o=a_i,\ x_t=a_k\}_{c_{rn}=\eta_{rn}}}.$$

In (81) the c_{rn} are arbitrary real numbers and the conditional expectation is regarded as a function of the c_{rn} evaluated at the (random) argument point $c_{rn} = \eta_{rn}$ $(r = 1, \ldots, n)$. From now on a normalizing factor will be denoted by the generic symbol N. Then

$$p_j^{(n)}(t) = N \sum_{i=1}^{K} p_i(0)p_{ij}(t)E\{\exp[\sum_{r=1}^{n} c_{rn}\xi_{rn}v_{rn}^{-1} - \frac{1}{2}\sum_{r=1}^{n}\xi_{rn}^2 v_{rn}^{-1}] | x_o = a_i,\ x_t = a_j\}_{c_{rn} = \eta_{rn}}. \tag{82}$$

By our assumptions on the $\tilde{x}(t)$ process, (82) can also be written

$$p_j^{(n)}(t) = N \sum_{i=1}^{K} p_i(0)p_{ij}(t)E\{\exp[\sum_{r=1}^{n} \eta_{rn}\tilde{\xi}_{rn}v_{rn}^{-1} - \frac{1}{2}\sum_{r=1}^{n}\tilde{\xi}_{rn}^2 v_{rn}^{-1}] | \tilde{x}_o = a_i,\ \tilde{x}_t = a_j,\ \eta_{1n}, \ldots, \eta_{nn}\}. \tag{83}$$

Since almost every $\tilde{x}(t)$ sample function is a step function, the limit

$$\lim_{n\to\infty} \sum_{r=1}^{n} \tilde{\xi}_{rn}^2 v_{rn} = \int_0^t \beta(s)^{-2}\tilde{x}(s)^2 ds \tag{84}$$

exists with probability 1 and hence, by dominated convergence, in mean square. Further

$$\sum_{r=1}^{n} \eta_{rn}\tilde{\xi}_{rn}v_{rn}^{-1} = \sum_{r=1}^{n} [\xi_{rn}\tilde{\xi}_{rn} + \tilde{\xi}_{rn}\int_{(r-1)t/n}^{rt/n} \beta(s)dw(s)]v_{rn}^{-1}$$

$$\to \int_0^t \beta(s)^{-2}x(s)\tilde{x}(s)ds + $$

$$+ \int_0^t \beta(s)^{-1}\tilde{x}(s)dw(s) \quad (n\to\infty)$$

$$= \int_0^t \beta(s)^{-2}\tilde{x}(s)dy(s), \tag{85}$$

where the integral with respect to w is a limit in mean square ([6] IX, §2). Let $\theta_n(t)$ be the random variable in [ ] in (83) and put

$$\theta(t) = \int_0^t \beta(s)^{-2}\tilde{x}(s)^2 ds - \frac{1}{2}\int_0^t \beta(s)^{-2}\tilde{x}(s)dy(s). \tag{86}$$

From (84) and (85), $\text{l.i.m.}\ \theta_n(t) = \theta(t)$. Furthermore

$$|e^{\theta_n(t)} - e^{\theta(t)}| \leqq \frac{1}{2}|\theta_n(t) - \theta(t)|[e^{\theta_n(t)} + e^{\theta(t)}]$$

$$\leqq |\theta_n(t) - \theta(t)|e^{\lambda|w(t)| + \mu t}$$

for some constants λ, $\mu > 0$; and since $E\{[e^{\lambda|w(t)|}]^2\} < \infty$,

$$\underset{n \to \infty}{\text{l.i.m.}}\; e^{\theta_n(t)} = e^{\theta(t)}. \tag{87}$$

Denote by $\mathcal{F}_t^n$ (resp. $\mathcal{F}_t$) the smallest Borel field relative to which the random variables $\tilde{x}(0)$, $\tilde{x}(t)$, $\eta_{1n}, \ldots, \eta_{nn}$, (resp. $\tilde{x}(0)$, $\tilde{x}(t)$, $y(s)$, $0 \leqq s \leqq t$) are measurable. Now

$$\begin{aligned} &E\{e^{\theta_n(t)}|\mathcal{F}_t^n\} - E\{e^{\theta(t)}|\mathcal{F}_t\} \\ &= E\{e^{\theta_n(t)}|\mathcal{F}_t^n\} - E\{e^{\theta_n(t)}|\mathcal{F}_t\} + \\ &\quad + E\{e^{\theta_n(t)}|\mathcal{F}_t\} - E\{e^{\theta(t)}|\mathcal{F}_t\}. \end{aligned} \tag{88}$$

Since $\mathcal{F}_t^n \subset \mathcal{F}_t$ and since, by inspection of $\theta_n(t)$, the random variable $E\{e^{\theta_n(t)}|\mathcal{F}_t\}$ is measurable relative to $\mathcal{F}_t^n$, there follows ([6] I, Thm. 8.1)

$$E\{e^{\theta_n(t)}|\mathcal{F}_t^n\} - E\{e^{\theta_n(t)}|\mathcal{F}_t\} = 0 \tag{89}$$

with probability 1. Finally, (87) implies

$$\underset{n \to \infty}{\text{l.i.m.}}\; E\{e^{\theta_n(t)}|\mathcal{F}_t\} - E\{e^{\theta(t)}|\mathcal{F}_t\} = 0. \tag{90}$$

Thus we have shown that

$$\underset{n\to\infty}{\text{l.i.m.}}\ p_j^{(n)}(t) = N \sum_{i=1}^{K} p_i(0)p_{ij}(t)E\{e^{\theta(t)} | \tilde{x}(0) = a_i,\ \tilde{x}(t) = a_j;\ y(s),\ 0 \leq s \leq t\}. \tag{91}$$

By exactly the same argument as in Appendix 1 the right side of (91) can be identified with $p_j(t)$.

2. Next we derive the Markov property. For $0 \leq \tau \leq t$ write

$$\varphi(\tau, t) = \exp[\int_\tau^t \beta(s)^{-2}\tilde{x}(s)^2 ds - \frac{1}{2}\int_\tau^t \beta(s)^{-2}\tilde{x}(s)dy(s)], \tag{92}$$

and let $\mathcal{H}_\tau^t$ be the smallest Borel field relative to which the random variables $y(s)$, $\tau \leq s \leq t$, are measurable. Thus $\varphi(\tau, t)$ is measurable relative to the Borel field generated by $\mathcal{H}_\tau^t$ and the $\tilde{x}(s)$'s for $\tau \leq s \leq t$. With this notation

$$p_j(t) = N \sum_{i=1}^{K} p_i(0)p_{ij}(t)E\{\varphi(0,t) | \tilde{x}_o = a_i,\ \tilde{x}_t = a_j,\ \mathcal{H}_o^t\}. \tag{93}$$

Now

$$\begin{aligned} p_j(t+h) &= N \sum_{k=1}^{K} p_k(0)p_{kj}(t+h)E\{\varphi(0,t+h) | \tilde{x}_o = a_k,\ \tilde{x}_{t+h} = a_j,\ \mathcal{H}_o^{t+h}\} \\ &= N \sum_{k=1}^{K} p_k(0)p_{kj}(t+h) \sum_{i=1}^{K} P\{\tilde{x}_t = a_i | \tilde{x}_o = a_k,\ \tilde{x}_{t+h} = a_j\} \\ &\qquad \cdot E\{\varphi(0,t)\varphi(t,t+h) | \tilde{x}_o = a_k,\ \tilde{x}_t = a_i,\ \tilde{x}_{t+h} = a_j, \mathcal{H}_t^{t+h}\} \\ &= N \sum_{i=1}^{K} \sum_{k=1}^{K} p_k(0)p_{ki}(t)p_{ij}(h)E\{\varphi(0,t) | \tilde{x}_o = a_k,\ \tilde{x}_t = a_i, \mathcal{H}_o^t\} \\ &\qquad \cdot E\{\varphi(t,t+h) | \tilde{x}_t = a_i,\ \tilde{x}_{t+h} = a_j,\ \mathcal{H}_t^{t+h}\}, \end{aligned} \tag{94}$$

where we have used the fact that the $\tilde{x}(t)$ process is Markov and is independent of the $y(t)$ process. Comparing (93) and (94) we obtain

$$p_j(t+h) = N \sum_{i=1}^{K} p_i(t) p_{ij}(h) E\{\varphi(t, t+h) | \tilde{x}_t = a_i, \tilde{x}_{t+h} = a_j, \mathcal{Y}_t^{t+h}\}. \tag{95}$$

Write $p(t) = [p_1(t), \ldots, p_k(t)]$ and consider the joint process $\{x(t), p(t), t \geq 0\}$. Eqs. (92) and (95) show that $x(t + h), p(t + h)$ depend only on $x(t), p(t)$ and on the $w(s)$ increments for $t \leq s \leq t + h$. Reasoning as on Appendix 2 we conclude that the joint process $\{x(t), p(t), t \geq 0\}$ is Markov.

3. We now evaluate functions m_i and b_{ij} defined by

$$m_j(t, x, p) = \lim_{h \to 0} E\{\frac{p_j(t+h) - p_j(t)}{h} \Big| x(t) = x, p(t) = p\} \tag{96}$$

and

$$b_{ij}(t,x,p) = \lim_{h \to 0} E\{\frac{[p_i(t+h) - p_i(t)][p_j(t+h) - p_j(t)]}{h} \Big| x(t) = x, \; p(t) = p\}. \tag{97}$$

To simplify computation note that the conditional expectation in (95) is readily evaluated, given the extra condition that no jump of $\tilde{x}(\cdot)$ occurs in the interval $(t, t + h)$; and the conditional probability that no jump occurs, given $\tilde{x}_t = \tilde{x}_{t+h} = a_i$, is

$$p_{ii}(h)^{-1} e^{-\nu_i h} = 1 + o(h) \quad (h \to 0). \tag{98}$$

Also, since $\beta(s)^{-1}$ is assumed bounded, we have

$$\varphi(t, t + h) \leq f(h) e^{\lambda |\Delta y|} \tag{99}$$

where $\Delta y = y(t+h) - y(t), \lambda > 0$ is constant and $f(h)$ is bounded as $h \to 0$. Put

$$\theta_j = a_j \int_t^{t+h} \beta(s)^{-2} dy(s) - \frac{1}{2} a_j^2 \int_t^{t+h} \beta(s)^{-2} ds. \tag{100}$$

From (92) and (98)-(100),

$$E\{\varphi(t,\ t+h)|\tilde{x}_t = a_i,\ \tilde{x}_{t+h} = a_i,\ \mathcal{H}_t^{t+h}\} = e^{\theta_i} + h\omega_i(h) \tag{101}$$

and if $j \neq i$

$$E\{\varphi(t,\ t+h)|\tilde{x}_t = a_i,\ \tilde{x}_{t+h} = a_j,\ \mathcal{H}_t^{t+h}\} = \omega_{ij}(h) \tag{102}$$

where (for a suitable $f(\cdot)$)

$$o \leqq \omega_i(h),\ \omega_{ij}(h) \leqq f(h)e^{\lambda|\Delta y|}. \tag{103}$$

Thus

$$p_j(t+h) = N[p_j(t)p_{jj}(h)\{e^{\theta_j} + h\omega_j(h)\} + \sum_{\substack{i \neq j \\ i=1}}^{K} p_i(t)p_{ij}(h)\omega_{ij}(h)]. \tag{104}$$

A simple calculation from (100) yields

$$\lim_{h \to 0} h^{-1}E\{\theta_j|x(t) = x\} = a_j x\beta(t)^{-2} - \frac{1}{2} a_j^2\beta(t)^{-2} \tag{105}$$

$$\lim_{h \to 0} h^{-1}E\{\theta_i\theta_j|x(t) = x\} = a_i a_j\beta(t)^{-2}. \tag{106}$$

Using (104)-(106) it is now straightforward to compute the limits (96) and (97). The results are

$$m_j(t,x,p) = -\nu_j p_j + \sum_{\substack{i=1 \\ i \neq j}}^{K} \nu_{ij}p_i + \beta(t)^{-2}(x - \bar{x})(a_j - \bar{x})p_j \tag{107}$$

and

$$b_{ij}(t,x,p) = [\beta(t)^{-1}(a_i - \bar{x})p_i][\beta(t)^{-1}(a_j - \bar{x})p_j] \tag{108}$$

where

$$\bar{x} = \sum_{i=1}^{K} a_i p_i . \tag{109}$$

4. Define

$$\sigma_j(t, x, p) = \beta(t)^{-1}(a_j - \bar{x})p_j, \quad j = 1, \ldots, K. \tag{110}$$

It will be shown that the $p(t)$ process can be represented as the solution of the stochastic differential system

$$dp_j(t) = m_j[t, x(t), p(t)]dt + \sigma_j[t, x(t), p(t)]dw(t); \qquad t \geqq 0, \quad j = 1, \ldots, K; \tag{111}$$

where $\{w(t), t \geq 0\}$ is the Wiener process introduced in (19). We shall apply a representation theorem of Doob ([6] VI, Thm. 3.3), generalized to allow for the fact that only the $p(t)$ component of the $\{x(t), p(t)\}$ process is continuous (that almost every $p(t)$ sample function is continuous follows from (104)).

The $p(t)$ process is obviously bounded; hence the conditions usually imposed on m and σ ([6] VI, §3) are satisfied. Now let $\mathcal{G}_t$ be the smallest Borel field with respect to which the random variables $x(s), p(s)$, $0 \leq s \leq t$ are measurable. Then, since the $\{x, p\}$ process is Markov, the evaluations (107), (108) are unchanged if the conditional expectations in (96), (97) are defined relative to $\mathcal{G}_t$. Reasoning as in the proof of [6] VI, Thm. 3.3, we conclude that each process

$$\{p_j(t) - p_j(0) - \int_0^t m_j[s, x(s), p(s)]ds, \ \mathcal{G}_t; \ t \geq 0\}, \ j=1, \ldots, K, \tag{112}$$

is a martingale which satisfies the conditions of [6] IX, Thm. 5.3. That is, if $\{\tilde{p}_j(t), \mathcal{G}_t; \ t \geq 0\}$ is the martingale defined by $(112)_j$, and if $\{\sigma_j\}^{-1} = \sigma_j^{-1}$ (or 0) where $\sigma_j \neq 0$ (or 0), then the equation

$$\hat{w}(t) = \int_0^t \{\sigma_j[s, x(s), p(s)]\}^{-1} d\tilde{p}_j(s) \tag{113}$$

defines a Wiener process; and by (108), (110) this definition of the $\hat{w}(t)$ process is independent of the choice of j.

It remains to identify the $\hat{w}(t)$ process with the $w(t)$ process of (19). Let

$$\begin{aligned} \tilde{y}(t) &= y(t) - \int_0^t x(s)ds \\ &= \int_0^t \beta(s)dw(s). \end{aligned} \tag{114}$$

Using (104) - (106) and (110) we find that

$$\begin{aligned} \lim_{h \to 0} h^{-1} E\{[\tilde{y}(t+h) - \tilde{y}(t)][p_j(t+h) - p_j(t)] \mid \mathcal{G}_t\} \\ = \beta(t)\sigma_j(t, x(t), p(t)]. \end{aligned} \tag{115}$$

Reasoning as before we conclude that the process $\{\tilde{y}(t), \mathcal{G}_t;\ t \geq 0\}$ is a martingale. Since

$$w(t) = \int_0^t \beta(s)^{-1} d\tilde{y}(s)$$

it follows by (113) and (115) that

$$E\{[w(t) - w(s)][\hat{w}(t) - \hat{w}(s)] \mid \mathcal{G}_s\} = t - s, \quad 0 < s < t. \tag{116}$$

Hence for each $t > 0$, $\hat{w}(t) = w(t)$ with probability 1; by continuity this implies that the processes $\hat{w}(t)$, $w(t)$ are essentially identical. The integrated form of (21) now follows from (113) by inversion.

References.

1. R. L. Stratonovich, Conditional Markov processes, Theory of Probability and its Applications 5 (2), 1960, pp. 156-178.

2. G. E. Kolosov and R. L. Stratonovich, A problem of synthesis of an optimal regulator by methods of dynamic programming, Avtomatika i Telemekhanika 24 (9), 1963, pp. 1165-1173.

3. W. M. Wonham, Stochastic Problems in Optimal Control, RIAS TR 63-14, May, 1963.

4. Investigation of Filter Functions, Final TR, Contract No. DA-36-034-ORD-3708Z, (Unclassified) U. S. Army Ordnance Missile Command, Redstone Arsenal, Alabama; January, 1964.

5. H. J. Kushner, On the differential equations satisfied by conditional probability densities of Markov processes, with applications (to be published); private communication, December, 1963.

6. J. L. Doob, Stochastic Processes, John Wiley and Sons, 1953.

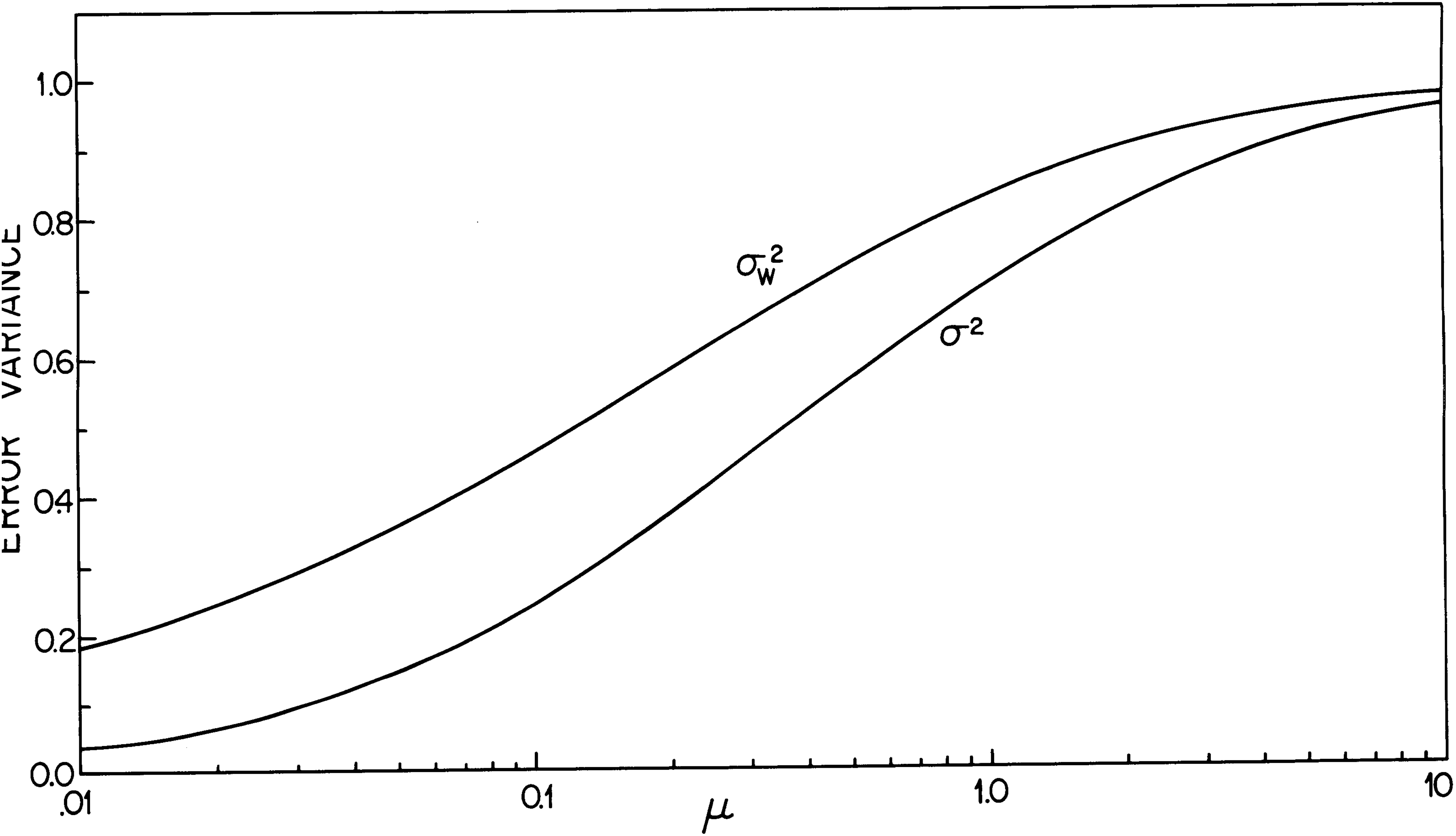
7. E. B. Dynkin, Markovskie Protsessi, Fizmatgiz, Moscow, 1963.

8. K. Ito, On Stochastic Differential Equations, Mem. Amer. Math. Soc., No. 4, 1951.

9. I. I. Gihman, On the theory of differential equations of stochastic processes, 1. Ukr. Math. J. 2 (4), 1950, pp. 37-63.

10. J. F. Barrett, Application of Kolmogorov's equations to randomly disturbed automatic control systems, Proc. IFAC, 1960 (Automatic and Remote Control, Butterworth, London, 1961) Vol. 2, pp. 724-733.

11. S.O. Rice, Mathematical analysis of random noise, Bell System Tech. J. 23, 24, 1944; reprinted in Noise and Stochastic Processes (ed. N. Wax), Dover Publ., 1954.


ERROR VARIANCE
1.0
0.8
0.6
0.4
0.2
0.0
.01
0.1
1.0
10
μ
σ_w^2
σ^2